

ORIE 4741 proposal: Curative effect prediction

Nanqing Dong (nd367), Ziyi Chen (zc286)

September 30, 2016

Our team aims at the curative effect prediction, and the selection of suitable hospitals (will be de-identified in our project), treatments, etc., for various diseases, which would be helpful for patients to recover with higher probability. We will adopt the dataset of 2544731 patients from

<https://health.data.ny.gov/Health/Hospital-Inpatient-Discharges-SPARCS-De-Identified/u4ud-w55t>

for the following reasons:

(1) In this dataset, the APR Risk of mortality is a good indicator of curative effect, so we will consider it as the target variable. In addition, for each patient, there are many relevant predictors such as the type of disease, the hospital, the treatment adopted, the length of stay in hospital, the severity level, etc.

(2) The dataset is big and messy for us to practice what we learnt from this course. There are missing values, categorical values, numerical values, etc.

To purely focus on the selection of a factor, we may attempt to control the other highly correlated features. For example, for each disease, while selecting the suitable hospitals, we may group the patients via the severity level of disease (categorical) and compare the hospitals group by group, since the hospitals with patients of higher severity level also tend to have higher risk of mortality. By controlling the severity level, and maybe the other highly correlated predictors, we can focus on the quality of hospitals only for a certain disease.